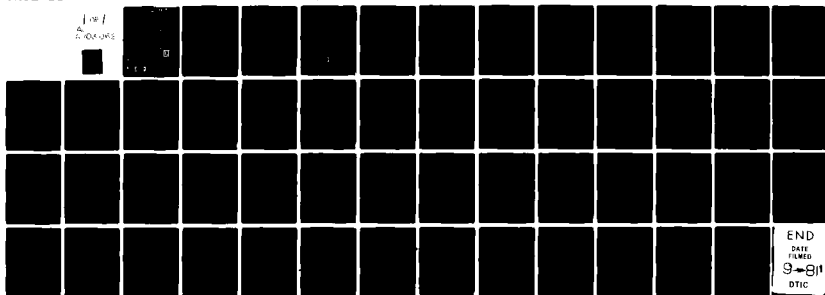


AD-A103 062

MASSACHUSETTS INST OF TECH LEXINGTON LINCOLN LAB F/8 9/3
THE MINIMAL DIVERGENCE SOLUTION TO THE GAUSSIAN MASKING PROBLEM--ETC(U)
MAY 81 P M FISHMAN, L K JONES, C W THERRIEN F19628-80-C-0002
TR-569 ESO-TR-81-96 NL

UNCLASSIFIED

10/1
A-100-0015



(12)
R

LEVEL II

See 1473

AD A103062

Technical Report

569 ✓

The Minimal Divergence Solution
to the Gaussian Masking ProblemP.M. Fishman
L.K. Jones
C.W. Therrien

22 May 1981

Prepared for the Department of the Air Force
under Electronic Systems Division Contract F19628-80-C-0002 by ✓**Lincoln Laboratory**
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LEXINGTON, MASSACHUSETTS

Approved for public release; distribution unlimited.

S DTIC
ELECTE D
AUG 1 8 1981
B

81 8 18 185

DTIC FILE COPY

The work reported in this document was performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology, with the support of the Department of the Air Force under Contract F19628-80-C-0002.

This report may be reproduced to satisfy needs of U.S. Government agencies.

The views and conclusions contained in this document are those of the contractor and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the United States Government.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

Raymond L. Loiselle

Raymond L. Loiselle, Lt. Col., USAF
Chief, ESD Lincoln Laboratory Project Office

Non-Lincoln Recipients

PLEASE DO NOT RETURN

Permission is given to destroy this document
when it is no longer needed.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LINCOLN LABORATORY

**THE MINIMAL DIVERGENCE SOLUTION
TO THE GAUSSIAN MASKING PROBLEM**

*P.M. FISHMAN
L.K. JONES
C.W. THERRIEN
Group 92*

TECHNICAL REPORT 569

22 MAY 1981

Approved for public release; distribution unlimited.

LEXINGTON

MASSACHUSETTS

i/ii

ABSTRACT

The problem of designing a stationary Gaussian noise process of fixed variance so as to optimally mask the possible presence of a given additive stationary Gaussian signal process is considered. A sub-optimal solution is obtained by minimizing the divergence distance between the noise and signal-plus-noise processes. Recursive time and frequency domain expressions for the divergence are derived in terms of successive auto-regressive approximations of the processes. For short observation times, the minimal divergence masking problem may then be solved by the unconstrained minimization of a convex - and recursively computable - function in the time domain. For long observation times, the problem reduces to that of minimizing the asymptotic divergence rate. This problem may be solved in the frequency domain by straight-forward algebraic techniques. A number of examples are given which illustrate the methodology.

DTIC
ELECTE
AUG 18 1981
S D B

Accession For

NTS ☒

PT ☐

U ☐

Dist ☐

By Codes

and/or

Dist ☐

A

CONTENTS

Abstract	iii
I. Introduction	1
II. Preliminary Definitions	4
III. Time and Frequency Domain Representations for the Directed Divergence	8
IV. The Minimal Divergence Masker	20
V. The Asymptotic Masking Problem	35
References	46

I. Introduction

In this report we consider some novel aspects of the problem of detecting a Gaussian process in additive Gaussian noise. Let the integer $N > 1$ be given and let a discrete stationary Gaussian process $s(t)$ (the signal process) be specified by its mean and covariance function on the interval $t = 1, 2, \dots, N$. We are to determine -- under a mean square energy constraint -- the statistics of the Gaussian noise process $m(t)$ (the masker process) on the interval $1, 2, \dots, N$ which maximizes an observers probability of error in the binary hypotheses test: masker process vs. masker process plus signal process. We refer to this problem as the Gaussian masking problem. The solution to this problem is of practical interest for the efficient design of jamming signals. It also enables a worst case analysis of detection performance for Gaussian signals where the noise spectrum is not specified in advance.

The explicit maximization of the probability of error expression for this problem is computationally burdensome, even for small N . As an alternative, we obtain the masker process which minimizes the divergence distance measure $[1,2]$ between the masker and masker plus signal processes. While this minimal divergence masker (MDM) solution is in

general sub-optimal, we emphasize it here for the following two reasons:

- (1) it offers the advantage of analytical and computational simplicity
- (2) it is shown in [3] that for large N and small signal/masker variance ratios, the probability of error, P_e , is given asymptotically by

$$P_e = \frac{1}{\sqrt{2\pi}} \int_{\sqrt{J}/2}^{\infty} \exp(-x^2/2) dx$$

where J is the divergence between the $m(t)$ and $m(t)+s(t)$ processes. Thus, the MDM solution corresponds to the optimal solution for this case.

Section II gives some definitions necessary for the succeeding development. In Section III, recursive time and frequency domain expressions for the divergence between arbitrary multivariate stationary Gaussian time series are given (Theorem 1). These expressions are in terms of the successive auto-regressive (AR) approximations (time domain) and corresponding AR spectral approximations (frequency domain) of the processes. Although our primary motivation for deriving these expressions is their applicability to the MDM solution of the masking problem, these expressions have intrinsic interest. The time domain expression is related to Akaike's information

criterion [4]. It is particularly convenient computationally since it may be evaluated efficiently using the multivariate Levinson algorithm [5]. The frequency domain expression is a generalization of expressions for the asymptotic divergence rate given in References 6-9. For completeness a new and simpler derivation of the latter expression based on the convergence properties of the AR spectral approximation is given in Theorem 2.

In Section IV, we apply the above representations to the MDM solution of the Gaussian masking problem. It is shown in Theorem 3 that for finite sample sizes ($N < \infty$) this problem may be solved in the time domain by the unconstrained minimization of a function of $N-1$ variables which has a unique local minimum (which is its global minimum). The MDM solution is then illustrated graphically for a number of numerical examples.

The asymptotic case ($N \rightarrow \infty$) is analyzed in Section V where it is shown that the MDM solution corresponds to the minimization of the asymptotic divergence rate between $m(t)$ and $m(t)+s(t)$. The unique solution to this problem is obtained in the frequency domain by straight-forward variational techniques (Theorem 4). It is also shown (Theorem 5) that the mean and discrete spectral component of $s(t)$ is irrelevant to the solution of the asymptotic masking problem.

II. Preliminary Definitions

Let \underline{X}_1 and \underline{X}_2 be two real random vectors with mutually absolutely continuous probability measures and respective density functions $f_1(\underline{X})$ and $f_2(\underline{X})$. Let $\Lambda_{1,2}(\underline{X}) = \log[f_1(\underline{X})/f_2(\underline{X})]$ denote the log-likelihood function. The directed divergences, $I(1,2)$ and $I(2,1)$ and the divergence, $J(1,2)$ are then defined by

$$\begin{aligned} I(1,2) &= E_1[\Lambda_{1,2}] \\ &= \int \cdots \int f_1(\underline{X}) \Lambda_{1,2}(\underline{X}) d\underline{X} \end{aligned} \quad (1.a)$$

$$I(2,1) = E_2[\Lambda_{2,1}] \quad (1.b)$$

$$J(1,2) = I(1,2) + I(2,1) \quad (2)$$

where $E_k[\cdot]$ denotes the expectation under hypotheses k .

These measures of statistical separability were first introduced by Jeffreys [10] and have found wide applicability to problems of feature extraction [1] and optimal signal design [2]. $I(1,2)$ is also referred to as the relative entropy or Kullback-Liebler Number of \underline{X}_1 with respect to \underline{X}_2 and may be interpreted as the mean information gained from the observation \underline{X}_1 for discrimination in favor of hypothesis 1.

For the case where the observations are Gaussian with respective mean vectors $\underline{\mu}_k$ and covariance matrices $\underline{\Sigma}_k$, $I(1,2)$ is given by ([1], p. 189).

$$I(1,2) = \frac{1}{2} \log \frac{|\underline{\Sigma}_2|}{|\underline{\Sigma}_1|} + \frac{1}{2} \text{tr } \underline{\Sigma}_1 (\underline{\Sigma}_2^{-1} - \underline{\Sigma}_1^{-1}) + \frac{1}{2} \text{tr } \underline{\Sigma}_2^{-1} (\underline{\mu}_1 - \underline{\mu}_2) (\underline{\mu}_1 - \underline{\mu}_2)^T. \quad (3)$$

For the case of interest here where the observations consist of N samples from one of two possible stationary Gaussian processes, recursive expressions are derived for $I(1,2)$ which are more computationally efficient than (3) and are of particular value in the context of the masking problem. These expressions involve AR approximations to the processes. Definitions of the requisite quantities are briefly summarized below.

Consider two M -variate discrete time stationary Gaussian processes $\{\underline{x}_{t,k}\}$ ($k = 1, 2$; $t=0, \pm 1, \pm 2, \dots$) with means \underline{M}_k and matrix covariance functions $\underline{C}_k(l) = E_k \left[(\underline{x}_{t+l,k} - \underline{M}_k) (\underline{x}_{t,k} - \underline{M}_k)^T \right]$. Let $\underline{\Sigma}_{n,k}$ ($n=1, 2, \dots$) be the block toeplitz matrix defined by

$$\underline{\Sigma}_{n,k} = \begin{bmatrix} \underline{C}_k(0) & \underline{C}_k^T(1) & \dots & \underline{C}_k^T(n-1) \\ \hline \underline{C}_k(1) & \underline{C}_k(0) & & \\ \hline \vdots & \vdots & \ddots & \\ \hline \underline{C}_k(n-1) & \underline{C}_k(n-2) & & \underline{C}_k(0) \\ \hline \end{bmatrix} \quad (4)$$

where $\underline{\Sigma}_{n,k}$ is positive definite for all n . Define $\tilde{\underline{\Sigma}}_{n,k}$ to be the matrix obtained by replacing the submatrix elements $\underline{C}_k(\ell)$ by $\tilde{\underline{C}}_k(\ell) = \underline{C}_k(\ell) + \underline{d}\underline{d}^T$ where $\underline{d} = \underline{M}_2 - \underline{M}_1$.

Let $\underline{S}_k(\lambda)$ denote the $M \times M$ power spectral matrix of process k , given by

$$\underline{S}_k(\lambda) = \frac{1}{2\pi} \sum_{\ell=-\infty}^{\infty} \underline{C}_k(\ell) e^{-i\lambda\ell}, \quad -\pi < \lambda < \pi. \quad (5)$$

Define $\tilde{\underline{S}}_k(\lambda)$ to be the spectral matrix of the process with covariance function $\tilde{\underline{C}}_k(\ell)$, i.e.

$$\tilde{\underline{S}}_k(\lambda) = \underline{S}_k(\lambda) + \underline{d}\underline{d}^T \delta(\lambda). \quad (6)$$

Let \underline{x}_t denote either process 1 or 2 and let

$$\underline{x}_n = \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \vdots \\ \underline{x}_n \end{bmatrix}$$

denote n successive observations of the process. Define the following random variables:

$$\hat{\underline{x}}_{n,k} = \begin{cases} \underline{M}_k & \text{for } n=1 \\ E_k [\underline{x}_n | \underline{x}_{n-1}] & \text{for } n=2,3,\dots \end{cases} \quad (7)$$

$$\underline{\varepsilon}_{n,k} = \underline{x}_n - \hat{\underline{x}}_{n,k}. \quad (8)$$

$\hat{\underline{x}}_{n,k}$ is the filtered estimate under hypothesis k of \underline{x}_n given the observation \underline{x}_{n-1} and $\underline{\varepsilon}_{n,k}$ is the corresponding error process. We also define $\underline{R}_{n,k} = E_k [\underline{\varepsilon}_{n,k} \underline{\varepsilon}_{n,k}^T]$ and $e_{n,k} = \text{tr } \underline{R}_{n,k}$ respectively as the n -th step error covariance and mean square error under hypothesis k .

The $(n-1)$ th order AR representation of the error process is then defined by

$$\sum_{j=0}^{n-1} b_{jn}^{(k)} (\underline{x}_{n-j} - \underline{M}_k) = \underline{\varepsilon}_{n,k} \quad (9)$$

$k=1,2$
 $n=1,2,3,\dots$

where $\{ \underline{b}_{j,n}^{(k)} \} \ (j=0,1,\dots,n-1)$ are the set of $(n-1)$ th order partial auto-regressive coefficient matrices of process k and $\underline{b}_{0,n}^{(k)} = \underline{I}$ without loss of generality. Define the $M \times Mn$ matrix $\underline{B}_{n,k}$ as

$$\underline{B}_{n,k} = \left[\underline{b}_{0,n}^{(k)} | \underline{b}_{1,n}^{(k)} | \dots | \underline{b}_{n-1,n}^{(k)} \right] .$$

Similarly the $(n-1)$ th order AR approximation of the power spectrum is defined by

$$\underline{S}_{n,k}(\lambda) = \frac{1}{2\pi} \left[\underline{\Gamma}_{n,k}(\lambda) \right]^{-1} \underline{R}_{n,k} \left[\underline{\Gamma}_{n,k}(\lambda)^* \right]^{-1}, \quad -\pi \leq \lambda < \pi \quad (10)$$

where

$$\underline{\Gamma}_{n,k}(\lambda) = \sum_{j=0}^{n-1} \underline{b}_{j,n}^{(k)} e^{-i\lambda \cdot j} \quad (11)$$

is the transfer function of the auto-regressive filter defined by (9). $\underline{S}_{n,k}(\lambda)$ may be shown to be the spectrum of the "most random" process whose covariance function matches that of process k on the lag interval $[-n+1, n-1]$ [11].

III. Time and Frequency Domain Representations for the Directed Divergence

Let $f_k(\underline{x}_N)$ denote the joint density of N successive observations under hypothesis k . The N -sample directed divergence, $I_N(1,2)$, and N -sample divergence, $J_N(1,2)$ are then

defined by analogy with (1) and (2) as

$$I_N(1,2) = E_1 \left[\Lambda_{1,2}(\underline{X}_N) \right] \quad (12)$$

$$J_N(1,2) = I_N(1,2) + I_N(2,1) . \quad (13)$$

The following theorem provides recursive formulas for the evaluation of (12) and (13).

Theorem 1. The N-sample directed divergence $I_N(1,2)$ has the following auto-regressive representations:

1. Time domain

$$I_N(1,2) = -\frac{MN}{2} + \frac{1}{2} \sum_{n=1}^N \left[\text{tr } \underline{B}_{n,2} \tilde{\underline{\Sigma}}_{n,1} \underline{B}_{n,2}^T \underline{R}_{n,2}^{-1} \right. \\ \left. - \log \frac{|\underline{R}_{n,1}|}{|\underline{R}_{n,2}|} \right] \quad (14)$$

2. Frequency domain

$$I_N(1,2) = -\frac{MN}{2} + \frac{1}{4\pi} \int_{-\pi}^{\pi} \sum_{n=1}^N \left[\text{tr}(\tilde{\underline{S}}_1(\lambda) \underline{S}_{n,2}^{-1}(\lambda)) \right. \\ \left. - \log \frac{|\underline{S}_{n,1}(\lambda)|}{|\underline{S}_{n,2}(\lambda)|} \right] d\lambda \quad (15)$$

Proof: We first derive (14). Let $f_k(\underline{x}_n | \underline{X}_{n-1})$ denote the conditional density of \underline{x}_n given \underline{X}_{n-1} under hypothesis k.

$\Lambda_{1,2}(\underline{X}_N)$ may be expressed recursively as

$$\Lambda_{1,2}(\underline{x}_n) = \Lambda_{1,2}(\underline{x}_{n-1}) + \log [f_1(\underline{x}_n|\underline{x}_{n-1})/f_2(\underline{x}_n|\underline{x}_{n-1})]$$

from which it follows that

$$I_n(1,2) = I_{n-1}(1,2) + \Delta_n, \quad n=1,2,\dots$$

where $I_0=0$ and

$$\Delta_n = \begin{cases} E_1[\log f_1(\underline{x}_1) - \log f_2(\underline{x}_1)] & , n=1 \\ E_1[\log f_1(\underline{x}_n|\underline{x}_{n-1}) - \log f_2(\underline{x}_n|\underline{x}_{n-1})] & , n>1 \end{cases}$$

Substituting

$$f_k(\underline{x}_n|\underline{x}_{n-1}) = (2\pi)^{-M/2} |\underline{R}_{n,k}|^{-1/2} \exp \left[-\frac{1}{2} \underline{\epsilon}_{n,k}^T \underline{R}_{n,k}^{-1} \underline{\epsilon}_{n,k} \right]$$

into the above gives

$$\begin{aligned} \Delta_n &= \frac{1}{2} E_1 \left[\underline{\epsilon}_{n,2}^T \underline{R}_{n,2}^{-1} \underline{\epsilon}_{n,2} - \underline{\epsilon}_{n,1}^T \underline{R}_{n,1}^{-1} \underline{\epsilon}_{n,1} \right. \\ &\quad \left. - \log (|\underline{R}_{n,1}|/|\underline{R}_{n,2}|) \right] = -\frac{1}{2} M - \frac{1}{2} \log (|\underline{R}_{n,1}|/|\underline{R}_{n,2}|) \\ &\quad + \frac{1}{2} E_1 \left[\underline{\epsilon}_{n,2}^T \underline{R}_{n,2}^{-1} \underline{\epsilon}_{n,2} \right] \end{aligned} \quad (16)$$

where the expectation appearing on the right of (16) may be evaluated as

$$\begin{aligned}
& E_1 \left[\begin{matrix} T & -1 \\ \underline{\varepsilon}_{n,2} & \underline{R}_{n,2} & \underline{\varepsilon}_{n,2} \end{matrix} \right] \\
& = E_1 \left[\left| \sum_{j=0}^{n-1} \underline{b}_{j,n}^{(2)} (\underline{x}_{n-j-\underline{M}_1}) - \left(\sum_{j=0}^{n-1} \underline{b}_{j,n}^{(2)} \right) \underline{d} \right|^T \underline{R}_{n,2}^{-1} \right. \\
& \quad \left. \left| \sum_{i=0}^{n-1} \underline{b}_{i,n}^{(2)} (\underline{x}_{n-i-\underline{M}_1}) - \left(\sum_{i=0}^{n-1} \underline{b}_{i,n}^{(2)} \right) \underline{d} \right| \right] \\
& = \text{tr} \left[\sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \underline{b}_{i,n}^{(2)} E_1 \left[(\underline{x}_{n-i-\underline{M}_1}) (\underline{x}_{n-j-\underline{M}_1})^T \right] \underline{b}_{j,n}^{(2)T} \underline{R}_{n,2}^{-1} \right] \\
& \quad + \text{tr} \left[\sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \underline{b}_{i,n}^{(2)} \underline{d} \underline{d}^T \underline{b}_{j,n}^{(2)T} \underline{R}_{n,2}^{-1} \right] \\
& = \text{tr} \left\{ \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \underline{b}_{i,n}^{(2)} \left[\underline{C}_1(j-i) + \underline{d} \underline{d}^T \right] \underline{b}_{j,n}^{(2)T} \underline{R}_{n,2}^{-1} \right\} \\
& = \text{tr} \underline{B}_{n,2} \tilde{\underline{\Sigma}}_{n,1}^T \underline{B}_{n,2}^{-1} .
\end{aligned}$$

The first part of the theorem, (14), follows from the latter result and (16).

We now derive (15). Let \underline{x}_t again represent either process 1 or 2 and denote its mean by \underline{M} and its spectrum by $\underline{S}(\lambda)$. Let k be fixed and define $\tilde{\underline{S}}(\lambda) = \underline{S}(\lambda) + (\underline{M} - \underline{M}_k)(\underline{M} - \underline{M}_k)^T \delta(\lambda)$. Let $n \geq 0$ be fixed and consider the process

$$\begin{aligned}
\underline{y}_t^{(k,n)} &= \sum_{j=0}^{n-1} \underline{b}_{j,n}^{(k)} (\underline{x}_{t-j-\underline{m}_k}) = \sum_{j=0}^{n-1} \underline{b}_{j,n}^{(k)} (\underline{x}_{t-j-\underline{m}-(\underline{m}_k-\underline{m})}) , \\
& \quad t=0, \pm 1, \pm 2, \dots .
\end{aligned}$$

The spectrum of $\underline{y}_t^{(k,n)}$ is given by

$$\underline{S}_y^{(k,n)}(\lambda) = \underline{\Gamma}_{n,k}(\lambda) \left[\underline{S}(\lambda) + (\underline{M} - \underline{M}_k)(\underline{M} - \underline{M}_k)^T \delta(\lambda) \right] \underline{\Gamma}_{n,k}^*(\lambda)$$

$$\underline{\Gamma}_{n,k}(\lambda) \underline{\tilde{S}}(\lambda) \underline{\Gamma}_{n,k}^*(\lambda).$$

Applying Parseval's theorem and (10) gives

$$\begin{aligned} E \left[\underline{y}_t^{(k,n)T} \underline{R}_{n,k}^{-1} \underline{y}_t^{(k,n)} \right] &= \text{tr} \underline{R}_{n,k}^{-1} E \left[\underline{y}_t^{(k,n)} \underline{y}_t^{(k,n)T} \right] \\ &= \text{tr} \int_{-\pi}^{\pi} \underline{R}_{n,k}^{-1} \underline{S}_y^{(k,n)}(\lambda) d\lambda \\ &= \text{tr} \int_{-\pi}^{\pi} \underline{R}_{n,k}^{-1} \underline{\Gamma}_{n,k}(\lambda) \underline{\tilde{S}}(\lambda) \underline{\Gamma}_{n,k}^*(\lambda) d\lambda \\ &= \text{tr} \int_{-\pi}^{\pi} \underline{R}_{n,k}^{-1} \underline{\Gamma}_{n,k}(\lambda) \left[\underline{S}_{n,k}(\lambda) \underline{S}_{n,k}^{-1}(\lambda) \underline{\tilde{S}}(\lambda) \right] \underline{\Gamma}_{n,k}^*(\lambda) d\lambda \\ &= \text{tr} \int_{-\pi}^{\pi} \left[\underline{R}_{n,k}^{-1} \underline{\Gamma}_{n,k}(\lambda) \underline{S}_{n,k}(\lambda) \underline{\Gamma}_{n,k}^*(\lambda) \right] \underline{\tilde{S}}(\lambda) \underline{S}_{n,k}^{-1}(\lambda) d\lambda \\ &= \frac{1}{2\pi} \text{tr} \int_{-\pi}^{\pi} \underline{\tilde{S}}(\lambda) \underline{S}_{n,k}^{-1}(\lambda) d\lambda. \end{aligned}$$

From (9) we have, letting $t=n$:

$$\underline{y}_n^{(k,n)} = \sum_{j=0}^{n-1} \underline{b}_{j,n}^{(k)} (\underline{x}_{n-j} - \underline{M}_k) = \underline{\epsilon}_{n,k}.$$

Letting $x_n = x_{n,1}$ and k be alternately 1 or 2 gives

$$E_1 \left[\begin{matrix} \varepsilon_{n1}^T & R_{n1}^{-1} & \varepsilon_{n1} \end{matrix} \right] = \frac{1}{2\pi} \text{tr} \int_{-\pi}^{\pi} \underline{S}_1(\lambda) \underline{S}_{n1}^{-1}(\lambda) d\lambda = M \quad (17)$$

$$E_1 \left[\begin{matrix} \varepsilon_{n2}^T & R_{n2}^{-1} & \varepsilon_{n2} \end{matrix} \right] = \frac{1}{2\pi} \text{tr} \int_{-\pi}^{\pi} \tilde{\underline{S}}_1(\lambda) \underline{S}_{n2}^{-1}(\lambda) d\lambda. \quad (18)$$

Substituting (18) into (16) and the expression for $I_N(1,2)$ gives

$$I_N(1,2) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \sum_{n=1}^N \left[\text{tr} \left(\tilde{\underline{S}}_1(\lambda) \underline{S}_{n2}^{-1}(\lambda) \right) - M - \log \left| \frac{R_{n1}}{R_{n2}} \right| \right] d\lambda.$$

To show that

$$\log |R_{ni}| = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log |\underline{S}_{ni}(\lambda)| d\lambda, \quad (19)$$

note that $\underline{S}_{n1}(\lambda)$ is the spectrum of the $(n-1)$ th order AR process \underline{z}_t defined by

$$\sum_{j=0}^{n-1} b_{jn}^{(1)} \underline{z}_{t-j} = \underline{\varepsilon}_t, \quad t=0, \pm 1, \pm 2, \dots \quad (20)$$

where $\underline{\varepsilon}_t$ is a sequence of identically distributed uncorrelated random variables with

$$E \left[\begin{matrix} \underline{\varepsilon}_t & \underline{\varepsilon}_t' \end{matrix} \right] = R_{n1}, \quad t=0, \pm 1, \pm 2, \dots$$

Let $\underline{\Sigma}_{my}$ denote the covariance matrix of m successive observations of \underline{z}_t and let R_{my} be the associated m th step error covariance. We have [12]

$$|\underline{R}_{my}| = |\underline{\Sigma}_{my}| / |\underline{\Sigma}_{m-1,y}| \quad m=1,2,\dots$$

However, from (20),

$$\underline{R}_{my} = \begin{cases} \underline{R}_{m1} & m < n \\ \underline{R}_{n1} & m \geq n \end{cases}$$

It follows from the above and a result in [13], pg. 66 that

$$\begin{aligned} \log |\underline{R}_{n1}| &= \lim_{m \rightarrow \infty} \log |\underline{R}_{my}| = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log |\underline{S}_Y(\lambda)| d\lambda \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \log |\underline{S}_{n1}(\lambda)| d\lambda \end{aligned}$$

which is (19).

We note that the relationship between the divergence and linear predictive filtering for arbitrary (not necessarily stationary) Gaussian processes was first shown by Schweppe [14]. Equation (14) may be considered to be a special case of Schweppe's result when applied to the stationary case.

The primary value of (14) is that it enables the efficient computation of $I_N(1,2)$ and $J_N(1,2)$ recursively by application of the Levinson algorithm which generates $\underline{B}_{n,i}$ and $\underline{R}_{n,i}$ in terms of $\underline{B}_{n-1,i}$, $\underline{R}_{n-1,i}$ and $\underline{C}_{n,i}$. The general multivariate version involves the solution

of both a forward and backward estimation problem and is described in [5,15]. Note that (14), when implemented with the Levinson algorithm, requires operations on $M \times M$ matrices. This compares with the direct evaluation of (3) which requires operations on $MN \times MN$ matrices.

The Asymptotic Case

For large N , more useful measures of statistical separation than (12) and (13) are the (asymptotic) entropy rate and divergence rate [6 - 9] defined respectively by

$$R_I(1,2) = \lim_{N \rightarrow \infty} I_N(1,2)/N \quad (21)$$

$$R_J(1,2) = \lim_{N \rightarrow \infty} J_N(1,2)/N. \quad (22)$$

For completeness we use the result (15) and the convergence property of the AR spectral approximation to obtain an alternative derivation of a theorem, first stated (without proof) by Pinsker [6]. A more complicated proof based on the triangular factorization of \sum_2^{-1} in (3) is provided in [7]. For convenience, we state and prove the theorem for the scalar case. The generalization to the multivariate case is straight-forward.

Theorem 2. Let $\{x_{t,1}\}$ and $\{x_{t,2}\}$ of Section II be scalar, zero-mean processes with respective spectral densities $S_1(\lambda)$ and $S_2(\lambda)$ where $S_1(\lambda)$ is bounded and $S_2(\lambda)$ is continuous and strictly positive on $-\pi \leq \lambda \leq \pi$. Then

$$R_I(1,2) = \Psi(1,2)$$

where

$$\Psi(1,2) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left[S_1(\lambda) S_2^{-1}(\lambda) - 1 - \log(S_1(\lambda) S_2^{-1}(\lambda)) \right] d\lambda. \quad (23)$$

The proof of (23) given below utilizes the result that if $\{g_n(x)\}$ is a sequence of non-negative uniformly integrable functions¹ on $[a,b]$ with $\lim_{n \rightarrow \infty} g_n(x) = g(x)$ (a.e.), then

$$\lim_{n \rightarrow \infty} \int_a^b g_n(x) dx = \int_a^b g(x) dx \quad (24)$$

Proof. From (15) we have

$$R_I(1,2) = \lim_{n \rightarrow \infty} \frac{1}{4\pi} \int_{-\pi}^{\pi} \left[S_1(\lambda) S_{n2}^{-1}(\lambda) - 1 - \log(S_{n1}(\lambda) S_{n2}^{-1}(\lambda)) \right] d\lambda.$$

For the logarithmic component of the above expression we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{4\pi} \int_{-\pi}^{\pi} \log S_{ni}(\lambda) d\lambda &= \frac{1}{2} \lim_{n \rightarrow \infty} \log R_{ni} \\ &= \frac{1}{2} \lim_{n \rightarrow \infty} \log \frac{|\Sigma_{ni}|}{|\Sigma_{n-1,i}|} = \frac{1}{2} \lim_{n \rightarrow \infty} \log \frac{|\Sigma_{ni}|}{n} \\ &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \log S_i(\lambda) d\lambda \end{aligned}$$

1 A sequence of non-negative functions $\{g_n(x)\}$ is said to be uniformly integrable on $[a,b]$ if for $\epsilon > 0$, there exists a $\delta > 0$ such that for all n , $\int_A g_n(x) dx < \epsilon$ for any set $A \subset [a,b]$ with measure $m(A) < \delta$.

where to obtain the last result we have again utilized the theorem on p. 66 of [13].

It remains to be shown that

$$\lim_{n \rightarrow \infty} \int_{-\pi}^{\pi} S_1(\lambda) S_{n2}^{-1}(\lambda) d\lambda = \int_{-\pi}^{\pi} S_1(\lambda) S_2^{-1}(\lambda) d\lambda. \quad (25)$$

We note that

$$\lim_{n \rightarrow \infty} S_2(\lambda) S_{n2}^{-1}(\lambda) = 1 \text{ for all } \lambda \quad (26)$$

and

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} S_2(\lambda) S_{n2}^{-1}(\lambda) d\lambda = 1 \text{ for all } n \quad (27)$$

where (26) follows from the convergence property of the AR spectral approximation and (27) is identical with (17).

Equations (26) and (27) imply that the sequence $g_n(\lambda) = S_2(\lambda)/S_{n2}(\lambda)$ is uniformly integrable on $[-\pi, \pi]$. To see this, let β be an arbitrary small positive number and let

$$E_n = \left\{ \lambda : |g_n(\lambda) - 1| < \beta \right\}.$$

It follows from (26) that $m(E_n) = 2\pi - \alpha_n$ where $\lim_{n \rightarrow \infty} \alpha_n = 0$.

For any set $A \subset [0, 2\pi]$ we then have

$$\begin{aligned}
\int_A g_n(\lambda) d\lambda &\leq \int_{A \cap E_n} g_n(\lambda) d\lambda + \int_{E_n^c} g_n(\lambda) d\lambda \\
&= \int_{A \cap E_n} g_n(\lambda) d\lambda + \int_{-\pi}^{\pi} g_n(\lambda) d\lambda - \int_{E_n} g_n(\lambda) d\lambda \\
&\leq (1+\beta)m(A) + 2\pi - (1-\beta)(2\pi - \alpha_n) \\
&\leq (1+\beta)m(A) + 2\pi\beta + \alpha_n.
\end{aligned}$$

Now, let the ϵ of the uniform integrability definition be given. It follows from the last inequality that there is an N such that for $n > N$, $\int_A g_n(\lambda) d\lambda < \epsilon$ if $m(A) < \delta^*$ where $\delta^* = (\epsilon - 2\pi\beta - \alpha_N)/(1+\beta)$. For each $n < N$, there will exist a δ_n such that $\int_A g_n(\lambda) d\lambda < \epsilon$ for all A with $m(A) < \delta_n$ ([16], p. 85). Choose δ in the uniform integrability definition to be $\delta = \min \{\delta^*, \delta_1, \delta_2, \dots, \delta_{N-1}\}$.

The uniform integrability of $\{g_n(\lambda)\}$ implies the uniform integrability of $S_1(\lambda)S_{n2}^{-1}(\lambda)$ since by the assumptions of Theorem 2 there exists a $C < \infty$ such that $S_1(\lambda)/S_2(\lambda) < C$ for $-\pi \leq \lambda \leq \pi$. Then

$$\begin{aligned}
\int_A S_1(\lambda)S_{n2}^{-1}(\lambda) d\lambda &= \int_A S_1(\lambda)S_2^{-1}(\lambda) g_n(\lambda) d\lambda \\
&\leq C \int_A g_n(\lambda) d\lambda.
\end{aligned}$$

(25) then follows. This completes the proof of Theorem 2.

The following corollary is an extension of (25) and is used in the treatment of the asymptotic masking problem in Section V.

Corollary. Let $S_1(\lambda)$ be bounded and let $S_2(\lambda)$ be of the form

$$S_2(\lambda) = S_{2,c}(\lambda) + \sum_{i=1}^q d_i^2 \delta(\lambda - \lambda_i)$$

where $S_{2,c}$ is continuous and strictly positive on $[0, 2\pi]$ and the sum on the right corresponds to a finite number of discrete contributions to the spectrum of power d_i^2 at the frequency λ_i ($i=1, 2, \dots, q < \infty$). Then

$$\lim_{n \rightarrow \infty} \int_{-\pi}^{\pi} S_1(\lambda) S_{n2}^{-1}(\lambda) d\lambda = \int_{-\pi}^{\pi} S_1(\lambda) S_{2,c}^{-1}(\lambda) d\lambda. \quad (28)$$

Proof. We have $S_1 S_{n2}^{-1} = S_1 S_{2,c}^{-1} g_{n,c}$ where $g_{n,c} = S_{2,c} S_{n2}^{-1}$. Now

$$\lim_{n \rightarrow \infty} g_{n,c} = 1 \quad (\text{a.e.}),$$

and

$$\begin{aligned} \lim_{n \rightarrow \infty} \int_{-\pi}^{\pi} g_{n,c} d\lambda &= \lim_{n \rightarrow \infty} \int_{-\pi}^{\pi} \left[S_{2,c}^{-1} - \sum_{i=1}^q d_i^2 \delta(\lambda - \lambda_i) \right] S_{n2}^{-1} d\lambda \\ &= 2\pi \lim_{n \rightarrow \infty} \left[1 - \sum_{i=1}^q d_i^2 S_{n2}^{-1}(\lambda_i) \right] = 2\pi. \end{aligned}$$

The identical argument used for the proof of (25) may now be repeated to show that $\{g_{nc}\}$, and thus $\{S_1 S_{n2}\}$, are uniformly integrable. It then follows that

$$\begin{aligned}
\lim_{n \rightarrow \infty} \int_{-\pi}^{\pi} s_1 s_{n2}^{-1} d\lambda &= \int_{-\pi}^{\pi} \lim_{n \rightarrow \infty} s_1 s_{n2}^{-1} d\lambda \\
&= \int_{-\pi}^{\pi} \left[s_1 s_{2c}^{-1} \lim_{n \rightarrow \infty} g_{n,c} \right] d\lambda = \int_{-\pi}^{\pi} s_1 s_{2c}^{-1} d\lambda .
\end{aligned}$$

IV. The Minimal Divergence Masker

In the following, we apply the results of Section III to the MDM solution of the Gaussian masking problem outlined in the Introduction. Specifically, let the signal process $s(t)$ be a given scalar stationary Gaussian process with mean μ_s and covariance function $C_s(\ell)$. We wish to hide the possible presence of $s(t)$ by imbedding it in another stationary Gaussian process $m(t)$ (the "masker process") which is independent of $s(t)$ and has a fixed mean μ_m and fixed variance V_m . The optimal solution of the masking problem is provided by that masker covariance function $C_m(\ell)$ on the lag interval $\ell = 1, 2, \dots, N-1$ which maximizes the probability of error P_e in the binary hypotheses test

$$H_1: \quad x(t) = m(t) \quad (\text{signal not present})$$

$$H_2: \quad x(t) = m(t) + s(t) \quad (\text{signal present})$$

$$t=1, 2, \dots, N$$

where $\{x(t); t=1, 2, \dots, N\}$ represents N observations from one of the two hypotheses and both hypotheses are considered equally likely.

Letting $\underline{X}_N = (x_1, x_2, \dots, x_N)^T$, $\underline{M}_{N,m}$ and $\underline{M}_{N,s}$ denote N -vectors all of whose components are μ_m and μ_s respectively,

and $\Lambda(\underline{X}_N) = \log p_2(\underline{X}_N)/p_1(\underline{X}_N)$ denote the log-likelihood ratio, we have

$$P_e = \frac{1}{2} \left[\int_{-\infty}^0 \Lambda f_2(\Lambda) d\Lambda + \int_0^{\infty} \Lambda f_1(\Lambda) d\Lambda \right] \quad (29)$$

where $f_i(\Lambda)$ is the density of Λ under hypotheses i and Λ is given by the quadratic form

$$\Lambda(\underline{X}_N) = \frac{1}{2} \log \frac{|\underline{\Sigma}_{N1}|}{|\underline{\Sigma}_{N2}|} - \frac{1}{2} \text{tr} \underline{\Sigma}_{N,2}^{-1} \left[\underline{X}_N - (\underline{M}_{N,m} + \underline{M}_{N,s}) \right] \\ \left[\underline{X}_N - (\underline{M}_{N,m} + \underline{M}_{N,s}) \right]^T + \frac{1}{2} \text{tr} \underline{\Sigma}_{N,1}^{-1} (\underline{X}_N - \underline{M}_{N,m}) (\underline{X}_N - \underline{M}_{N,s})^T$$

where $\underline{\Sigma}_{N,1} = \underline{\Sigma}_{N,m}$, $\underline{\Sigma}_{N,2} = \underline{\Sigma}_{N,m} + \underline{\Sigma}_{N,s}$ and $\underline{\Sigma}_{N,m}$ and $\underline{\Sigma}_{N,s}$ are the analogues of (4).

The complicated nature of the densities $f_i(\Lambda)$ make the minimization of P_e with respect to $C_m(\ell)$ ($\ell=1,2,\dots,N-1$) numerically difficult. A sub-optimal solution to this problem may be obtained by determining that masking covariance which minimizes the divergence $J_N(1,2) = E_2(\Lambda) - E_1(\Lambda)$ between the masking and masking plus signal processes. This minimization -- involving only first moments of Λ -- is considerably simpler than that required for (29). However, the possibly complex nature of the $(N-1)$ dimensional surface of $J_N(1,2)$ (considered as a function of $C_M(1), C_M(2), \dots, C_M(N-1)$) may still make the minimization difficult - particularly for large N . Fortunately, the latter issue is of little concern

as the function has only one local minimum (which is the unique global minimum). This is shown in the following theorem which is valid for arbitrary Gaussian vectors.

Theorem 3. Let \underline{Y}_m (the masker) and \underline{Y}_s (the signal) be two independent N-dimensional Gaussian vectors with means $\underline{\alpha}_m$ and $\underline{\alpha}_s$ and positive definite covariance matrices \underline{K}_m and \underline{K}_s where $\underline{\alpha}_m$, $\underline{\alpha}_s$ and \underline{K}_s are fixed and \underline{K}_m is allowed to vary. Let $\underline{Y}_1 = \underline{Y}_m$ and $\underline{Y}_2 = \underline{Y}_m + \underline{Y}_s$. Then the divergence $J(1,2) \triangleq J(\underline{K}_m)$ between \underline{Y}_1 and \underline{Y}_2 is a strictly convex function of \underline{K}_m , i.e. for $0 < \theta < 1$ and arbitrary covariances \underline{K}_m , \underline{K}'_m with $\underline{K}_m \neq \underline{K}'_m$:

$$J(\theta \underline{K}_m + (1-\theta) \underline{K}'_m) < \theta J(\underline{K}_m) + (1-\theta) J(\underline{K}'_m) \quad (30)$$

Proof: The proof of (30) involves the simultaneous diagonalization of a judicious choice of two matrices associated with \underline{K}_m , \underline{K}'_m and \underline{K}_s . Let \underline{K}_m and \underline{K}'_m be fixed and consider the function

$$D(\theta) = J(\theta \underline{K}_m + (1-\theta) \underline{K}'_m), \quad 0 < \theta < 1.$$

Our approach will be to show that $d^2 D(\theta)/d\theta^2 > 0$ for all $\theta \in (0,1)$. We have from (2) and (3) that

$$\begin{aligned}
J(\underline{K}_m) &= \frac{1}{2} \underline{\alpha}_S^T \left[(\underline{K}_m + \underline{K}_S)^{-1} + \underline{K}_m^{-1} \right] \underline{\alpha}_S \\
&+ \frac{1}{2} \text{tr} \left[(\underline{K}_m + \underline{K}_S)^{-1} \underline{K}_m + \underline{K}_m^{-1} (\underline{K}_m + \underline{K}_S) - 2\underline{I} \right] \\
&= \frac{1}{2} \underline{\alpha}_S^T \left[(\underline{K}_m + \underline{K}_S)^{-1} + \underline{K}_m^{-1} \right] \underline{\alpha}_S \\
&+ \frac{1}{2} \text{tr} \left[\underline{K}_m^{-1} \underline{K}_S - (\underline{K}_m + \underline{K}_S)^{-1} \underline{K}_S \right] .
\end{aligned}$$

Thus

$$D(\theta) = D_1(\theta) + D_2(\theta)$$

where

$$\begin{aligned}
D_1(\theta) &= \frac{1}{2} \underline{\alpha}_S^T \left[(\underline{H}(\theta) + \underline{K}_S)^{-1} + \underline{H}^{-1}(\theta) \right] \underline{\alpha}_S \\
D_2(\theta) &= \frac{1}{2} \text{tr} \left[\underline{H}^{-1}(\theta) \underline{K}_S - (\underline{H}(\theta) + \underline{K}_S)^{-1} \underline{K}_S \right]
\end{aligned}$$

and

$$\underline{H}(\theta) = \theta \underline{K}_m + (1-\theta) \underline{K}'_m .$$

Let θ_0 with $0 < \theta_0 < 1$ be fixed and apply a non-singular linear transformation $\underline{T} = \underline{T}(\theta_0)$ such that

$$\underline{T} (\underline{H}(\theta_0)) \underline{T}^T = \underline{\Lambda} \tag{31}$$

$$\underline{T} \underline{K}_S \underline{T}^T = \underline{I} \tag{32}$$

where $\underline{\Lambda}$ is a diagonal matrix. For arbitrary θ , let $\bar{D}(\theta)$, $\bar{D}_1(\theta)$, $\bar{D}_2(\theta)$, $\bar{H}(\theta)$, $\bar{\alpha}_S$ and \bar{K}_S denote $D(\theta)$, $D_1(\theta)$, $D_2(\theta)$, $H(\theta)$,

$\underline{\alpha}_s$ and \underline{K}_s respectively, under the transformation $T(\theta_0)$.

From the invariance of the divergence to non-singular linear transformations it follows that

$$\bar{D}(\theta) = D(\theta) \quad , \quad 0 < \theta < 1$$

and in particular that

$$d^2 \bar{D}/d\theta^2 = d^2(\bar{D}_1 + \bar{D}_2)/d\theta^2 = d^2 D/d\theta^2 \quad (33)$$

$$0 < \theta < 1$$

Evaluating $d^2 \bar{D}_1/d\theta^2$ by repeated application of the inverse matrix differentiation rule gives

$$d^2 \bar{D}_1/d\theta^2 = \underline{V}^T (\underline{H} + \underline{K}_s)^{-1} \underline{V} + \underline{U}^T \underline{H}^{-1} \underline{U}$$

where

$$\underline{V} = (\underline{K}_m - \underline{K}_m') (\underline{H} + \underline{K}_s)^{-1} \underline{\alpha}_s$$

$$\underline{U} = (\underline{K}_m - \underline{K}_m') \underline{H}^{-1} \underline{\alpha}_s .$$

Since \underline{H} and $\underline{H} + \underline{K}_s$ are symmetric and positive definite, it follows that \underline{H}^{-1} and $(\underline{H} + \underline{K}_s)^{-1}$ are symmetric and positive definite. Thus

$$d^2 \bar{D}_1/d\theta^2 \geq 0 \quad , \quad 0 < \theta < 1 . \quad (34)$$

Similarly, differentiation of \bar{D}_2 gives

$$d^2 \bar{D}_2/d\theta^2 = \text{tr} \left\{ (\underline{K}_m - \underline{K}_m')^2 \underline{K}_s \left[\underline{H}^{-3} - (\underline{H} + \underline{K}_s)^{-3} \right] \right\}$$

Evaluating the latter expression at θ_0 and using (31) and (32) gives

$$d^2 \bar{D}_2/d\theta^2|_{\theta=\theta_0} = \text{tr} (\bar{\underline{K}}_m - \bar{\underline{K}}'_m)^2 \underline{R} \quad (35)$$

where $\underline{R} = \underline{\Lambda}^{-3} - (\underline{I} + \underline{\Lambda})^{-3}$ is a diagonal, positive definite matrix and $(\bar{\underline{K}}_m - \bar{\underline{K}}'_m)^2$ is non-negative definite. We now show that

$$d^2 \bar{D}_2/d\theta^2|_{\theta=\theta_0} > 0 . \quad (36)$$

Evaluating (35) gives

$$d^2 \bar{D}_2/d\theta^2|_{\theta=\theta_0} = \sum_{i=1}^N \beta_{ii} v_{ii}$$

where β_{ii} and v_{ii} are the i -th diagonal elements of $(\bar{\underline{K}}_m - \bar{\underline{K}}'_m)^2$ and \underline{R} respectively.

Since $\beta_{ii} = \sum_e (\bar{k}_{ie} - \bar{k}'_{ie})^2 \geq 0$ where $\bar{k}_{i,e}$ and $\bar{k}'_{i,e}$ are the elements of $\bar{\underline{K}}_m$, $\bar{\underline{K}}'_m$ respectively, and by hypotheses $\bar{\underline{K}}_m \neq \bar{\underline{K}}'_m$, it follows that at least one of the $\beta_{ii} > 0$ and (36) follows.

We conclude from (33), (34) and (36) that

$$d^2 D/d\theta^2|_{\theta=\theta_0} > 0 . \quad (37)$$

Since θ_0 is arbitrary, (37) is true for all θ with $0 < \theta < 1$.

Now consider the function $g(\theta) = D(\theta) - \theta D(1) - (1-\theta)D(0)$.

We have $g''(\theta) > 0$ for $0 < \theta < 1$, $g(\theta)$ continuous for $0 \leq \theta \leq 1$, and $g(0) = g(1) = 0$. It follows from these properties that $g(\theta) < 0$ for $0 < \theta < 1$, or, equivalently,

$$D(\theta) < \theta D(0) + (1-\theta)D(1)$$

which is (30). This completes the proof of Theorem 3.

We now return to the MDM problem as formulated for the stationary case.

We note that the divergence, $J_N(1,2)$ between the $m(t)$ and $m(t) + s(t)$ processes may be expressed using (2) and (14) as

$$J_N(1,2) = -N + \frac{1}{2} \sum_{n=1}^N \left[\underline{B}_{n;m+s} \tilde{\underline{\Sigma}}_{nm} \underline{B}'_{n;m+s} e_{n;m+s}^{-1} + \underline{B}_{n;m} \left(\underline{\Sigma}_{nm} + \tilde{\underline{\Sigma}}_{ns} \right) \underline{B}'_{n;m} e_{n;m}^{-1} \right]$$

where: $\tilde{\underline{\Sigma}}_{nm}$ and $\tilde{\underline{\Sigma}}_{ns}$ are obtained by adding μ_s^2 to each element of $\underline{\Sigma}_{nm}$ and $\underline{\Sigma}_{ns}$ respectively; $\underline{B}_{n;m}$ and $e_{n;m}$ are the $(n-1)$ th order AR parameters associated with $m(t)$; and $\underline{B}_{n;m+s}$ and $e_{n;m+s}$ are the $(n-1)$ th order AR parameters associated with $m(t) + s(t)$.

From the Yule Walker equations [17]

$$\underline{B}_{n;m} \underline{\Sigma}_{n;m} \underline{B}'_{n;m} = e_{n;m}$$

and the above expression simplifies to the following symmetric form

$$J_N(1,2) = \frac{1}{2} \left\{ -N + \sum_{n=1}^N \left[\underline{B}_{n;m+s} \sum_{nm} \underline{B}'_{n;m+s} e^{-1}_{n;m+s} + \underline{B}_{n;m} \sum_{ns} \underline{B}'_{nm} e^{-1}_{n;m} \right] \right\} . \quad (38)$$

The MDM process is then specified by the $N-1$ quantities $\hat{C}_m(1), \hat{C}_m(2), \dots, \hat{C}_m(N-1)$ which minimize (38). If this minimization is performed directly on the covariances $\{C_m(l)\}$ it is a problem of constrained optimization in that the solution must satisfy the positive definiteness constraint

$$\sum_{i=1}^N \sum_{j=1}^N \ell_i \ell_j \hat{C}_m(i-j) > 0 \quad (39)$$

for any set of numbers $(\ell_1, \ell_2, \dots, \ell_N)$ not all 0.

Fortunately there exists an alternative parameterization of $m(t)$ on the interval $t=1,2,\dots,N$ in terms of which the minimization of (38) is a problem of unconstrained minimization.

This is the parameterization in terms of $(\rho_{1,m}, \rho_{2,m}, \dots, \rho_{N-1,m})$ where $\rho_{j,m}$ is the j -th partial correlation coefficient of $m(t)$. The major advantage to the latter parameterization is that, in contrast to (39), the stationarity constraint on the $\rho_{j,m}$ is considerably simpler and is given by [17]

$$|\rho_{j,m}| < 1 \quad j = 1, 2, \dots, N-1 \quad . \quad (40)$$

(38) may be computed recursively in terms of the $\{\rho_{j,m}\}$ using the Levinson recursion. It may also be shown that the solution $(\hat{\rho}_{1,m}, \dots, \hat{\rho}_{N-1,m})$ which minimizes (38) in the closed $(N-1)$ -dimensional cube $|\rho_{j,m}| \leq 1$ ($n=1, 2, \dots, N-1$) is in the interior of the cube. Since the mapping from $(c_m(1), \dots, c_m(N-1))$ to $(\rho_{1,m}, \dots, \rho_{N-1,m})$ is one-to-one and continuous, it follows from Theorem 3 that (38) considered as a function of $(\rho_{1,m}, \dots, \rho_{N-1,m})$ has a unique local minimum (which is the global minimum) in the region specified by (40). This minimum may be obtained by standard numerical techniques for unconstrained minimization.

Numerical Examples.

We illustrate the MDM solutions corresponding to three different signal processes, $s(t)$. These are:

(a) $s(t) = s_1(t)$ where $s_1(t)$ is a first order Gauss-Markov process with mean μ_s and unit variance, satisfying the difference equation

$$s_1(t) = \mu_s + \frac{1}{2} [s_1(t-1) - \mu_s] + \epsilon_1(t) \quad (41)$$

$$t=0, \pm 1, \pm 2, \dots$$

where $\epsilon_1(t)$ is a zero-mean white Gaussian sequence.

(b) $s(t) = s_2(t)$ where

$$s_2(t) = s_1(t) + A \cos(2\pi f t) + B \sin(2\pi f t) \quad (42)$$

$$t=0, \pm 1, \pm 2, \dots$$

where $f=1/4$ and A and B are zero-mean uncorrelated Gaussian random variables with $\text{Var } A = \text{Var } B = 1/8$.

(c) $s(t) = s_3(t)$ where $s_3(t)$ is a Gaussian second-order auto-regressive process with mean μ_s and unit variance satisfying the difference equation

$$s_3(t) = \mu_s + 3/4 [s_3(t-1) - \mu_s] - 1/4 [s_3(t-2) - \mu_s] + \epsilon_3(t) \quad (43)$$

$$t=0, \pm 1, \pm 2, \dots$$

where $\epsilon_3(t)$ is a zero-mean white Gaussian sequence.

In the following, we illustrate the corresponding MDM solutions for various values of μ_s , masking variance (V_m), and sample size (N). Rather than give these solutions in the time domain (for instance by giving $\{\hat{C}_m(\ell), \ell=1, 2, \dots, N-1\}$, it is generally more informative to provide them in the frequency domain in terms of the corresponding (N-1)th order AR spectral approximations

$$\hat{s}_{N,m}(\lambda) = \frac{1}{2\pi} \frac{\hat{e}_{N,m}}{\left| \sum_{j=0}^{N-1} \hat{b}_{j,N}^{(m)} e^{-i\lambda j} \right|^2} \quad (44)$$

where $\hat{e}_{N,m}$ and $\{\hat{b}_{j,N}^{(m)}; j=0,1,\dots,N-1\}$ are the AR parameters corresponding to $\{C_m(0) = V_m, \hat{C}_m(1), \dots, \hat{C}_m(N-1)\}$.

Figures 1-3 show the MDM solutions corresponding to the three illustrative signal processes assuming $\mu_s = 0$, $N=20$ and $V_m = 10$ (Figs. 1-2) or $V_m = 3$ (Fig. 3) (the ripples in Fig. 2 are caused by ringing inherent to the AR spectral approximation). We note that the MDM spectra are not identical to the signal spectra and involve a shift of spectral energy relative to the signal from high to low energy regions. We refer to this as a "partial whitening" of the MDM. Intuitively, such an energy allocation is plausible since it allows the masker to dominate the signal in frequency regions where the signal is low with little decrease in the masker/signal ratio in regions where the signal is high.

Figure 4 shows the effect on the MDM of introducing a non-zero signal mean μ_s . As μ_s increases, more masking energy is concentrated in an impulsive manner in the dc region of the spectrum. This impulse corresponds in the time domain to the mean of $X_m(t)$ being a random variable (with a variance given by the strength of the impulse). This result has an obvious intuitive explanation: if, for instance, $\mu_s > 0$ and the N-sample observation has a high sample mean (e.g., $> \mu_m$) the observer must decide whether this is due to the presence

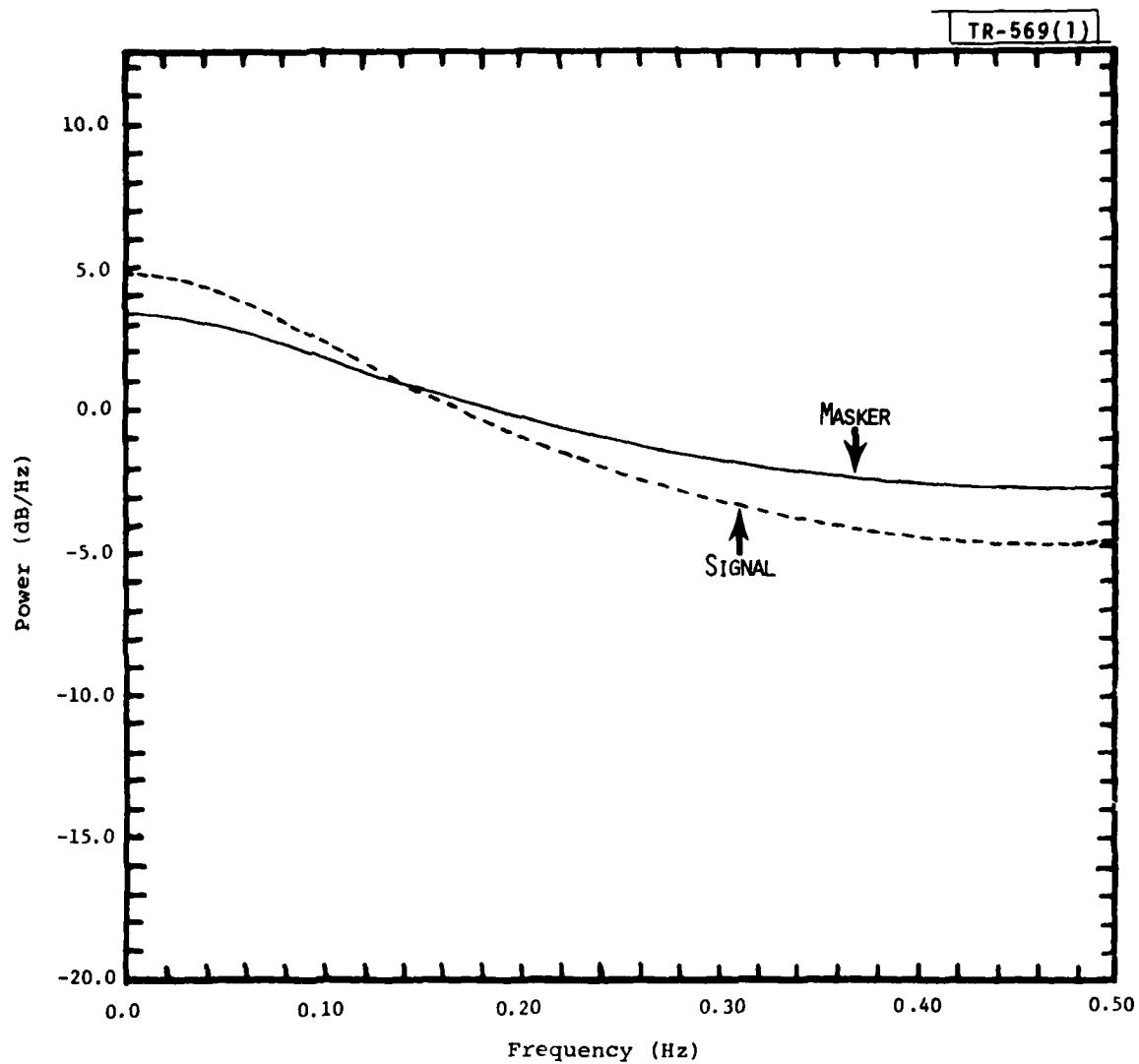


Fig. 1. Masker spectrum corresponding to zero-mean signal process defined by equation (41).

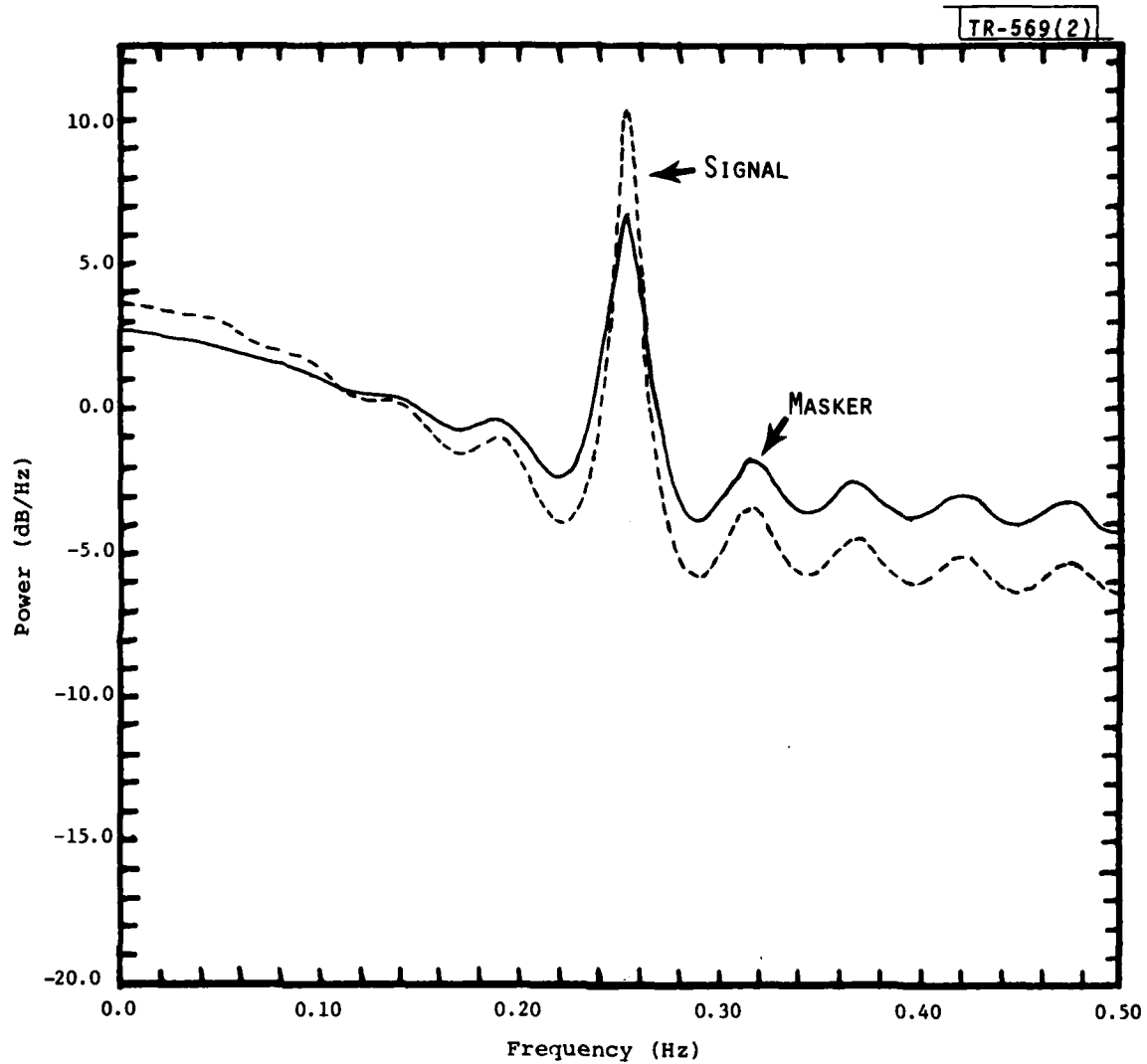


Fig. 2. Masker spectrum corresponding to zero-mean signal process defined by equation (42).

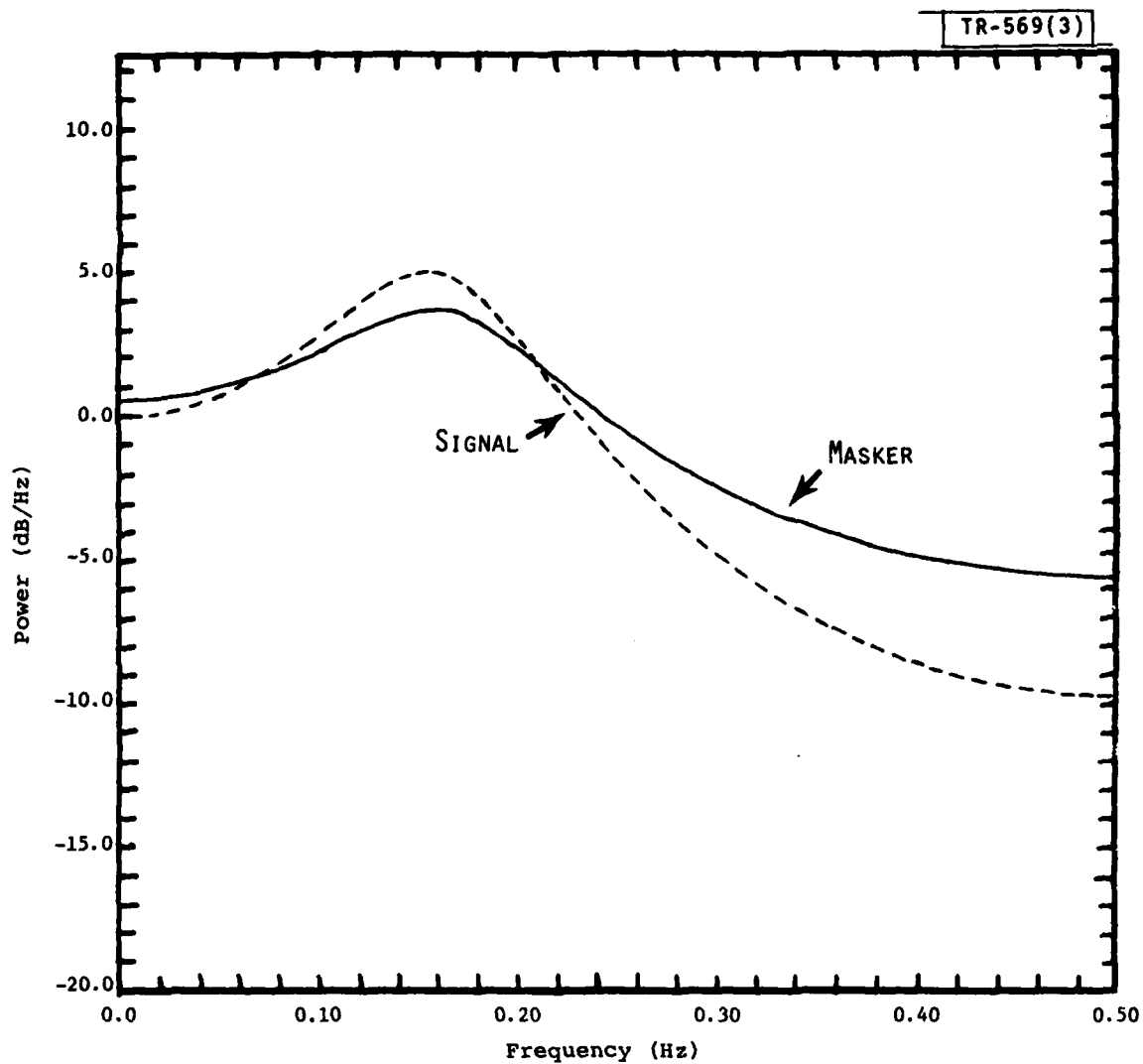


Fig. 3. Masker spectrum corresponding to zero-mean signal process defined by equation (43).

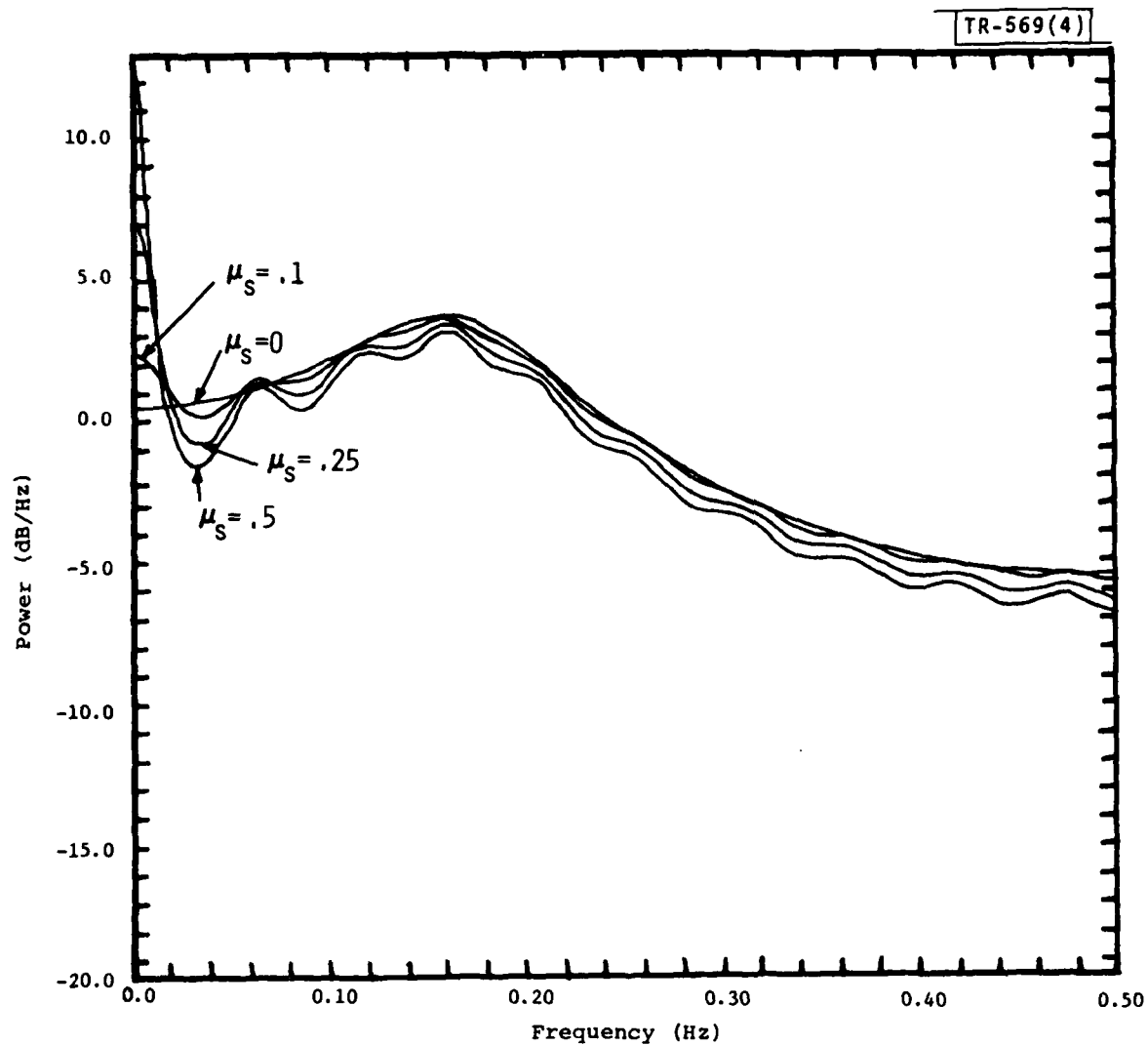


Fig. 4. Masker spectra corresponding to signal process defined by equation (43) with $\mu_s = 0.0, 0.1, 0.25$ and 0.5 .

of the signal or to the absence of the signal and the circumstance that the (random) mean of $X_m(t)$ for this sample was particularly high; conversely, if the sample mean is low, this may be due to the absence of the signal or to the presence of the signal and a mean of $X_m(t)$ which is particularly low.

Figures 5 and 6 show the effect of increasing N on the MDM for $s(t) = s_1(t)$ and $\mu_s = \frac{1}{2}$. Note (Fig. 5) that the energy in the DC region decreases as N goes from 10 to 40. Figure 6 shows the optimal masking auto-correlation $\{\hat{R}_m(\ell); \ell=0,1,\dots, N-1\}$ for $N = 5, 10, 20$, and 40. These results suggest that asymptotically ($N \rightarrow \infty$), the MDM solution for $\mu_s \neq 0$ converges to that for $\mu_s = 0$. That this is indeed the case in general (with a similar result valid for the discrete component of the signal spectrum), is shown in the next section, which treats the asymptotic masking problem.

V. The Asymptotic Masking Problem

Asymptotically ($N \rightarrow \infty$) the MDM problem becomes one of determining the $m(t)$ which minimizes the divergence rate $R_J(1,2)$ between the $m(t)$ and $m(t) + s(t)$ processes. We refer to the masking process which minimizes $R_J(1,2)$ as the asymptotic minimal divergence masker (AMDM). Our starting point for the analysis of the asymptotic case is the frequency domain analog

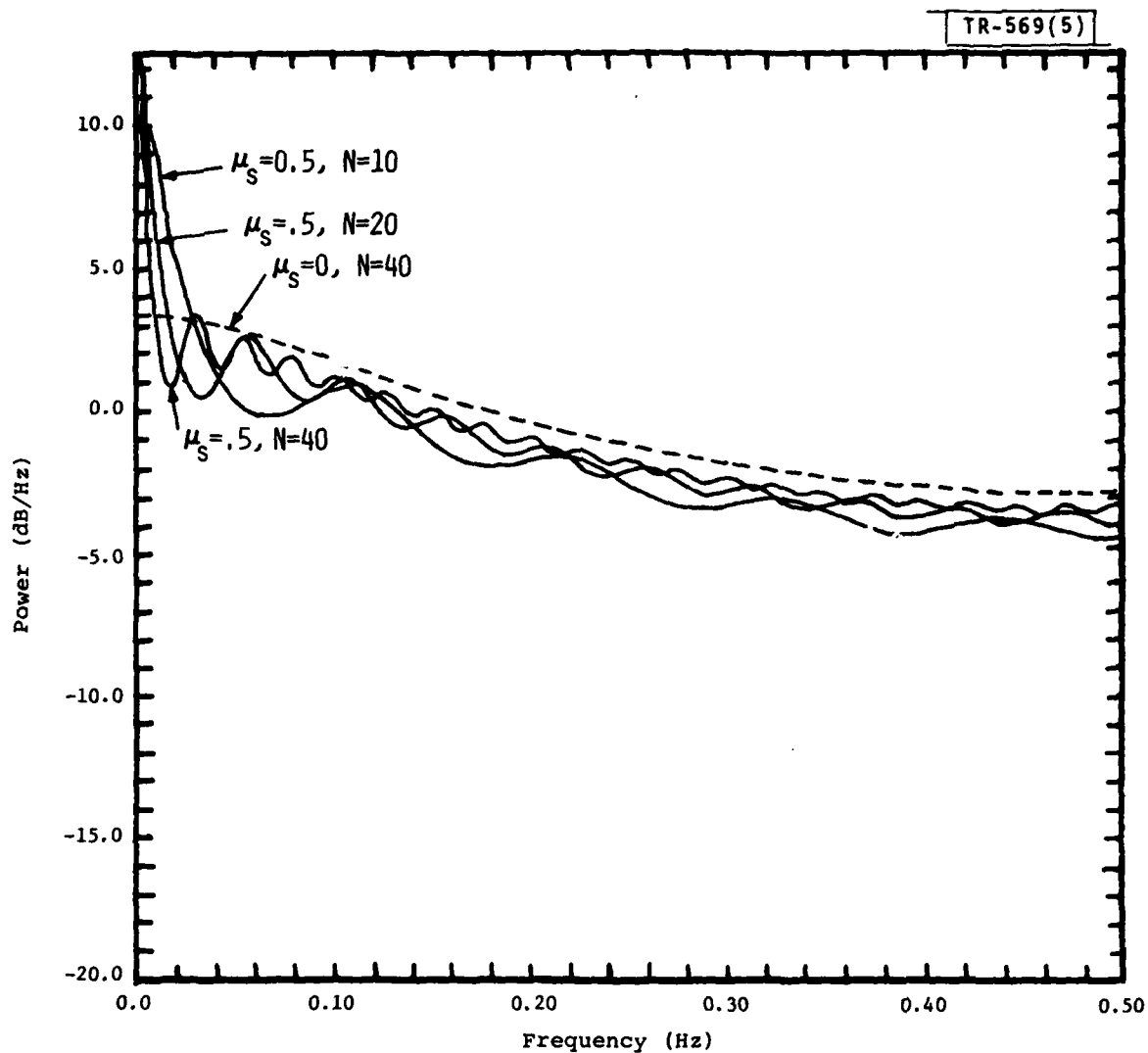


Fig. 5. Masker spectra corresponding to signal process defined by equation (41) with $\mu_s = 0.5$ and $N = 10, 20, 40$.

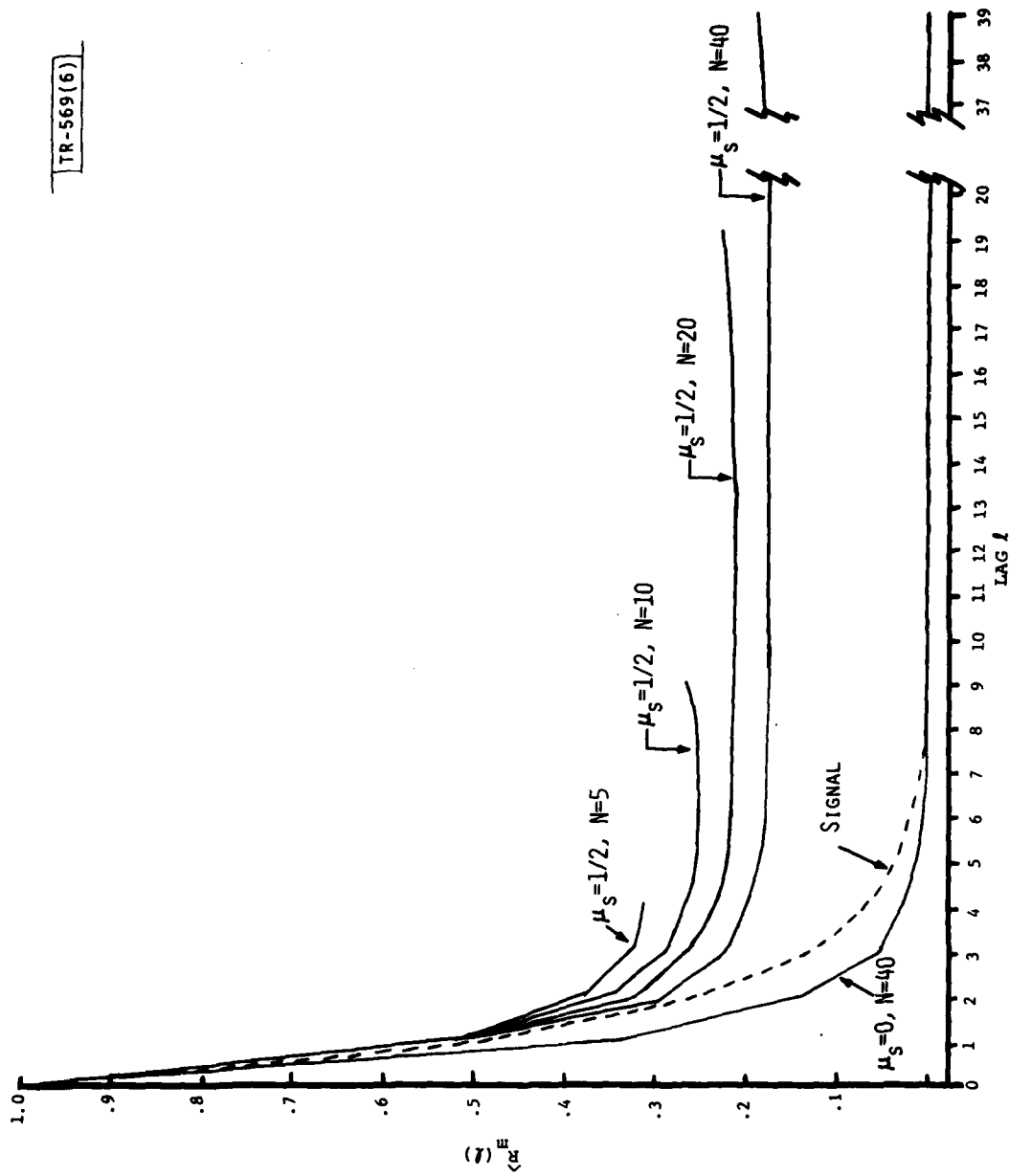


Fig. 6. Auto-correlation functions of masker processes corresponding to signal process defined by equation (41) with $\mu_s = 0.5$ and $N = 5, 10, 20, 40$.

of (38) which is obtained using (15) and (17) as

$$J_N(1,2) = \frac{1}{4\pi} \sum_{n=1}^N \int_{-\pi}^{\pi} \left[\frac{\tilde{S}_m}{S_{n;m+s}} - \frac{S_m}{S_{n;m}} + \frac{\tilde{S}_{m+s}}{S_{n;m}} - \frac{S_{m+s}}{S_{n;m+s}} \right] d\lambda \quad (45)$$

where: S_m and S_s are the spectra of X_m and X_s , $S_{m+s} = S_m + S_s$,

$S_{n;m}$ and $S_{n;m+s}$ are the $(n-1)$ th order AR approximations to S_m and S_{m+s} , and $\tilde{S}_m = S_m + \mu_s^2 \delta(\lambda)$, $\tilde{S}_{m+s} = S_m + S_s + \mu_s^2 \delta(\lambda)$.

Substituting the expressions for \tilde{S}_m and \tilde{S}_{m+s} into (45) gives

$$J_N(1,2) = \frac{1}{4\pi} \sum_{n=1}^N \left\{ \frac{\mu_s^2}{S_{n;m}(0)} + \frac{\mu_s^2}{S_{n;m+s}(0)} + \int_{-\pi}^{\pi} \left[\frac{S_s}{S_{n;m}} - \frac{S_s}{S_{n;m+s}} \right] d\lambda \right\} \quad (46)$$

Consider the case where $\mu_s \neq 0$ and the signal spectrum is of the form

$$S_s(\lambda) = S_{s,c}(\lambda) + \sum_{i=1}^q d_i^2 \delta(\lambda - \lambda_i) \quad (47)$$

where $S_{s,c}(\lambda)$ denotes the portion of the spectrum continuous on $[0, 2\pi]$ and the sum in (47) corresponds to a finite number of discrete contributions to the spectrum of power d_i^2 at the frequency λ_i ($i=1, 2, \dots, q < \infty$). (46) then becomes

$$J_N(1,2) = \frac{1}{4\pi} \sum_{n=1}^N \left\{ \frac{\mu_s^2}{S_{n;m}(0)} + \frac{\mu_s^2}{S_{n,m+s}(0)} + \sum_{i=1}^q \left[\frac{d_i^2}{S_{n,m}(\lambda_i)} - \frac{d_i^2}{S_{n,m+s}(\lambda_i)} \right] + \int_{-\pi}^{\pi} \left[\frac{S_{s,c}(\lambda)}{S_{n,m}(\lambda)} - \frac{S_{s,c}(\lambda)}{S_{n,m+s}(\lambda)} \right] d\lambda \right\}. \quad (48)$$

This is the general frequency domain expression for $J_N(1,2)$ in the context of the masking problem. The MDM masking solutions illustrated in Figs. 1-6 are best understood by inspection of (48).

The following two theorems characterize the AMDM solution for processes of the form (47) and provide a constructive method for its generation numerically.

Theorem 4. Let the signal process be zero-mean and have a continuous spectral density $S_s(\lambda) = S_{s,c}(\lambda) > 0$. Then in the class M of continuous positive spectral densities with variance V_m , there is a unique corresponding AMDM process. Its spectrum \hat{S}_m and divergence rate $\hat{R}_J(1,2)$ satisfy

$$\hat{R}_J(1,2) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left[\frac{S_s}{\hat{S}_m} - \frac{S_s}{\hat{S}_m + S_s} \right] d\lambda = \inf_{S_m \in M} \frac{1}{4\pi} \int_{-\pi}^{\pi} \left[\frac{S_s}{S_m} - \frac{S_s}{S_m + S_s} \right] d\lambda. \quad (49)$$

\hat{S}_m may be obtained to any desired approximation by a straightforward numerical procedure.

Proof. That the AMDM solution - if it exists - satisfies

(49) follows from the result that for any $S_m \in M$

$$\begin{aligned} R_J(1,2) &= \lim_{n \rightarrow \infty} \frac{1}{4\pi} \int_{-\pi}^{\pi} \left[\frac{S_s}{S_{n;m}} - \frac{S_s}{S_{n;m+s}} \right] d\lambda \\ &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \left[\frac{S_s}{S_m} - \frac{S_s}{S_m + S_s} \right] d\lambda \end{aligned} \quad (50)$$

which follows from (48) and (25). We now show that there is a unique $\hat{S}_m \in M$ by the explicit minimization of (50).

To minimize (50), we introduce the Lagrange multiplier ℓ and the functional (of S_m and ℓ)

$$\int_{-\pi}^{\pi} \left[\frac{S_s}{S_m} - \frac{S_s}{S_m + S_s} \right] d\lambda - \ell \left(\int_{-\pi}^{\pi} S_m d\lambda - v_m \right). \quad (51)$$

The critical points of (51) satisfy

$$\int_{-\pi}^{\pi} S_m d\lambda = v_m \quad (52)$$

and also satisfy for each λ :

$$S_s \left[(S_s + S_m)^{-2} - S_m^{-2} \right] = \ell$$

or

$$-\ell^{-1} (S_s^3 + 2 S_m S_s^2) = S_m^4 + 2 S_m^3 S_s + S_m^2 S_s^2. \quad (53)$$

ℓ must be < 0 since S_s and S_m are > 0 . We now show that for fixed $\ell < 0$ and fixed $\lambda \in [-\pi, \pi]$, there is a unique positive solution $S_{m;\ell}(\lambda)$ of (53).

Consider the graphs of

$$h_\ell(y) = -\ell^{-1} S_s^3 - 2 \ell^{-1} S_s^2 y$$

and

$$g(y) = y^4 + 2 S_s y^3 + S_s^2 y^2$$

for $y > 0$. Since $S_s > 0$, the slope of h_ℓ is a positive constant and $h_\ell(0) > 0$. Now $g(0) = 0 < h_\ell(0)$. Let \tilde{y}_ℓ be the smallest positive abscissa where g and h_ℓ intersect. Clearly, there are intersections since g grows faster than h_ℓ . Now $g'(y)$ is an increasing function and $g'(\tilde{y}_\ell)$ is necessarily greater than $h'_\ell(\tilde{y}_\ell)$ (= a constant). Since $g' > h'_\ell$ for $y > \tilde{y}_\ell$, $g > h_\ell$ for $y > \tilde{y}_\ell$. Hence, for fixed ℓ , g and h_ℓ intersect at exactly one point in the right half plane, and there is a unique $S_{m;\ell}(\lambda) = \tilde{y}_\ell$ which satisfies (53) for each λ . Moreover, it follows from the assumed continuity of $S_s(\lambda)$ in (53) that $S_{m;\ell}(\lambda)$ is a continuous function of λ .

In the above manner we have constructed by standard algebraic methods the family of continuous spectral densities $S_{m;\ell}(\lambda)$, parameterized by ℓ , which satisfy the constraint (53). We now determine the unique $\ell = \hat{\ell}$ such that $S_{m;\hat{\ell}} = \hat{S}_m$ also satisfies the constraint (52). Note that for $0 > \ell_2 > \ell_1$, $\tilde{y}_{\ell_2} > \tilde{y}_{\ell_1}$. Hence $S_{m;\ell_2}(\lambda) > S_{m;\ell_1}(\lambda)$ for all λ , and

$$\int_{-\pi}^{\pi} S_{m;\ell_2} d\lambda > \int_{-\pi}^{\pi} S_{m;\ell_1} d\lambda .$$

Thus the $\hat{\lambda}$ for which $S_{m;\hat{\lambda}}$ satisfies (52) is unique. $\hat{\lambda}$ may be found by a straight-forward binary search: start with λ_1 and $\lambda_2 > \lambda_1$ such that

$$\int_{-\pi}^{\pi} S_{m;\lambda_2} d\lambda > v_m, \int_{-\pi}^{\pi} S_{m;\lambda_1} d\lambda < v_m.$$

Determine λ_3 and λ_4 such that

$$\int_{-\pi}^{\pi} S_{m;\lambda_3} d\lambda < v_m, \int_{-\pi}^{\pi} S_{m;\lambda_4} d\lambda > v_m$$

and $\lambda_4 - \lambda_3 = \frac{1}{2}(\lambda_2 - \lambda_1)$, . . . and so on. The sequence of intervals $[\lambda_{2n+1}, \lambda_{2n+2}]$ converges to $\hat{\lambda}$. The search process may be terminated when an $\bar{\lambda}$ is obtained for which $\int_{-\pi}^{\pi} S_{m;\bar{\lambda}} d\lambda \approx v_m$ to any desired degree of approximation. $S_{m;\bar{\lambda}}$ is then the desired approximation to \hat{S}_m . This completes the proof of Theorem 4.

Corollary. $\hat{R}_J(1,2)$ is a monotonically decreasing function of $\gamma = v_m/v_s$.

Proof. Let $p_s(\lambda) = S_s(\lambda)/v_s$ and $p_m(\lambda) = S_m(\lambda)/v_m$ denote the normalized signal and masker spectral densities. We have

$$\begin{aligned} \int_{-\pi}^{\pi} \left[\frac{S_s}{S_m} - \frac{S_s}{S_m + S_s} \right] d\lambda &= \int_{-\pi}^{\pi} \left[\frac{p_s}{\gamma p_m} - \frac{p_s}{\gamma p_m + p_s} \right] d\lambda \\ &= \int_{-\pi}^{\pi} \frac{p_s^2}{\gamma p_m (\gamma p_m + p_s)} d\lambda. \end{aligned}$$

Since the last expression is a monotonically decreasing function of γ , the corollary then follows from (49).

Theorem 5. Let the signal process have mean $\mu_s \neq 0$ and a spectrum

$$S_s(\lambda) = S_{s,c}(\lambda) + \sum_{i=1}^q d_i^2 \delta(\lambda - \lambda_i) \quad (54)$$

as in (47). Then the corresponding AMDM process is in the class M and is the AMDM process corresponding to the zero-mean signal process with spectrum $S_{s,c}(\lambda)$ (i.e. the mean of the signal and the discrete component of its spectrum do not effect its AMDM).

Proof. Consider masker processes of the form

$$S_m(\lambda) = S_{m,c}(\lambda) + \epsilon_0^2 \delta(\lambda) + \sum_{i=1}^q \epsilon_i^2 \delta(\lambda - \lambda_i) \quad (55)$$

where $S_{m,c}$ is the continuous spectral component, $\epsilon_0^2 \delta(\lambda)$ is a discrete component corresponding to $\mu_s \neq 0$, and the sum on the right represents discrete components corresponding to the discrete components of S_s . Let $V_{m,d} = \sum_{i=0}^q \epsilon_i^2$ and $V_{m,c} = V_m - V_{m,d}$ be the partitioning of the masker variance between the discrete and continuous portions of the spectrum. Assuming that each of the ϵ_i^2 are >0 , we have for the corresponding AR approximations

$$\begin{aligned}
\lim_{n \rightarrow \infty} \hat{S}_{n,m}(0) &= \lim_{n \rightarrow \infty} \hat{S}_{n,m}(\lambda_i) \\
&= \lim_{n \rightarrow \infty} \hat{S}_{n,m+s}(\lambda_i) = \infty \quad . \quad (56) \\
&\quad i=1,2,\dots,q
\end{aligned}$$

From (48), (56) and (28) we have

$$\begin{aligned}
R_J(1,2) &= \lim_{N \rightarrow \infty} J_N(1,2) - J_{N-1}(1,2) \\
&= \frac{1}{4\pi} \lim_{n \rightarrow \infty} \int_{-\pi}^{\pi} \left[\frac{S_{s,c}}{S_{n,m}} - \frac{S_{s,c}}{S_{n,m+s}} \right] d\lambda \\
&= \frac{1}{4\pi} \int_{-\pi}^{\pi} \left[\frac{S_{s,c}}{S_{m,c}} - \frac{S_{s,c}}{S_{m,c}+S_{s,c}} \right] d\lambda \quad . \quad (57)
\end{aligned}$$

For fixed $V_{m,c} = V_m - \sum_{i=0}^q \epsilon_i^2$, the AMDM masker is then

fully characterized by Theorem 4 and is obtained by minimizing (57) with respect to the continuous component $S_{m,c}$ under the constraint $\int_{-\pi}^{\pi} S_{m,c} = V_{m,c}$. Denote the corresponding minimal divergence rate by $\hat{R}_J(1,2;V_{m,c})$. However, by the corollary to Theorem 4, $\hat{R}_J(1,2;V_{m,c})$ is a monotonically decreasing function of $V_{m,c}$. Thus for the AMDM solution, $V_{m,c} = V_m$, $V_{m,d} = 0$ and the theorem follows.

Remark - The fact that the AMDM masker spectrum has no discrete components even if such components are present in the signal, is best motivated by the following observation [6]:

consider two Gaussian processes with spectra

$$S_i(\lambda) = S_{i,c}(\lambda) + \sum_{j=1}^q a_{i,j}^2 \delta(\lambda - \lambda_j) \quad , i=1,2$$

where the $a_{i,j}^2 > 0$ for $i=1,2; j=1,2,\dots,q < \infty$. Then the asymptotic ($N \rightarrow \infty$) probability of error for distinguishing between the two processes is > 0 if and only if $S_{1,c}(\lambda) = S_{2,c}(\lambda)$ (a.e.). The fact that asymptotically non-singular detection depends on the equality of S_{1c} and S_{2c} but not on the equality of the $\{a_{1j}\}$ and $\{a_{2j}\}$, suggests that the AMDM place asymptotically vanishing energy in the discrete spectral components so as to maximize the similarity of the masker and masker plus signal continuous spectral components, as is indicated by Theorem 5.

REFERENCES

1. S. Kullback, Information Theory and Statistics, (Wiley, New York, 1959).
2. T. Kailath, "The Divergence and Bhattacharyya Distance Measures in Signal Selection," IEEE Trans. Commun. COM-15, 52 (1967).
3. L. K. Jones, "Some Remarks on the Gaussian Discriminant," submitted to IEEE Trans. Inf. Theory.
4. H. Akaike, "Canonical Correlation of Time Series and the Use of an Information Criterion," in System Identification: Advances and Case Studies, R. K. Mehra and D. G. Lainotis, Eds. (Academic Press, New York, 1976).
5. A. Lindquist, "A New Algorithm for Optimal Filtering of Discrete-Time Stationary Processes," SIAM J. Control 12, 736 (1974).
6. M. S. Pinsker, Information and Information Stability of Random Variables and Processes (Holden-Day, San Francisco, 1964).
7. R. H. Shumway and A. N. Unger, "Linear Discriminant Functions for Stationary Time Series," J. Am. Stat. Assoc. 69, 948 (1974).
8. R. M. Hawkes and J. B. Moore, "Performance Bounds for Adaptive Estimation," Proc. IEEE 64, 1143 (1976).
9. A. M. Yaglom, "On the Equivalence or Perpendicularity of Two Gaussian Probability Measures in Function Space," in Symposium on Time Series Analysis, edited by M. Rosenblatt (Wiley, New York, 1963).
10. H. Jeffreys, "An Invariant Form for the Prior Probability in Estimation Problems," Proc. R. Soc. A 186, 453 (1946).

11. J. A. Edward and M. M. Fitelson, "Notes on Maximum Entropy Processing," in Modern Spectrum Analysis, edited by D. G. Childers (IEEE, New York, 1978).
12. T. W. Anderson, The Statistical Analysis of Time Series (Wiley, New York, 1971).
13. V. Grenander and G. Szego, Toeplitz Forms and Their Applications (Univ. California, Berkely, 1958).
14. F. C. Schweppe, "On the Bhattacharyya Distance and the Divergence Between Gaussian Processes," Inform. and Control 11, 373 (1967).
15. C. W. Therrien, "A Sequential Approach to Target Discrimination," IEEE Trans, Aerospace Electron. Systems, AES-14, 433 (1978).
16. H. L. Royden, Real Analysis (MacMillan, New York, 1967).
17. G. E. P. Box and G. M. Jenkins, Time Series Analysis Forecasting and Control (Holden-Day, San Francisco, 1970).

12 53

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM	
1. REPORT NUMBER (7) ESD-TR-81-96	2. GOVT ACCESSION NO. AD A103062	3. RECIPIENT'S CATALOG NUMBER (9)	
4. TITLE (and Subtitle) (6) The Minimal Divergence Solution to the Gaussian Masking Problem		5. TYPE OF REPORT & PERIOD COVERED Technical Report	
7. AUTHOR(s) (2) Philip M. Fishman, Lee K. Jones and Charles W. Therrien		6. PERFORMING ORG. REPORT NUMBER Technical Report 569	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Lincoln Laboratory, M.I.T. P.O. Box 73 Lexington, MA 02173		8. CONTRACT OR GRANT NUMBER(s) (15) F19628-80-C-0002	
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Systems Command, USAF Andrews AFB Washington, DC 20331		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Program Element No. 63311F Project No. 627A	
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Electronic Systems Division Hanscom AFB Bedford, MA 01731 (16) 627A		12. REPORT DATE (11) 22 May 1981	
		13. NUMBER OF PAGES 54	
		15. SECURITY CLASS. (of this report) Unclassified	
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited. (14) TR-569		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE	
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)			
18. SUPPLEMENTARY NOTES None			
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Gaussian processes autoregressive modeling stationary Gaussian signal processes divergence for Gaussian autoregressive processes Gaussian masking problem			
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The problem of designing a stationary Gaussian noise process of fixed variance so as to optimally mask the possible presence of a given additive stationary Gaussian signal process is considered. A sub-optimal solution is obtained by minimizing the divergence distance between the noise and signal-plus-noise processes. Recursive time and frequency domain expressions for the divergence are derived in terms of successive auto-regressive approximations of the processes. For short observation times, the minimal divergence masking problem may then be solved by the unconstrained minimization of a convex - and recursively computable - function in the time domain. For long observation times, the problem reduces to that of minimizing the asymptotic divergence rate. This problem may be solved in the frequency domain by straight-forward algebraic techniques. A number of examples are given which illustrate the methodology.			

END

DATE
FILMED

9-81

DTIC